

The Evolution of Consensus in Macroeconomic Forecasting

October 2002
first version July 2001

Allan W. Gregory and James Yetman*

Abstract

When professional forecasters repeatedly forecast macroeconomic variables, their forecasts may converge over time towards a consensus. The evolution of consensus is analyzed with Blue Chip data under a parametric polynomial decay function that permits flexibility in the decay path. For the most part, this specification fits the data. We test whether forecast differences decay to zero at the same point in time for a panel of forecasters, and discuss possible explanations for this, along with its implications for studies using panels of forecasters.

JEL codes: C23 E37

* Gregory: Department of Economics, Queen's University, Kingston Ontario Canada K7L 3N6; Yetman: School of Economics and Finance, University of Hong Kong, Pokfulam Road, Hong Kong. For advice and assistance with data, we are grateful to Carole Constant, Robert Eggert, and Randall Moore. We wish to thank Associate Editor, Fred Joutz and two referees for helpful comments. Also, Bryan Campbell, Russell Davidson, and Gregor Smith provided useful criticisms on an earlier version of this paper.

1. Introduction

Decision makers in both the private and public sectors typically have panels of macroeconomic forecasts available to them. In such panels, it is common for forecasters to update their forecasts repeatedly. The degree of divergence or agreement among the forecasters, and changes in this over the forecast horizon, may serve as an important indicator of both forecaster behaviour and the degree of uncertainty in the variable being forecast.

Consider a panel of hypothetical forecasters, forecasting the growth rate of gross domestic product for 1998. Starting many months prior to the end of 1998, forecasters publish predictions of the growth rate. These initial predictions are updated repeatedly, and over time are likely to become increasingly similar for at least three reasons. First, forecasters observe the predictions of others, and are able to incorporate these into their own later revisions. Second, much of the new information that is incorporated into later revisions is common to all forecasters. And third, once we enter the 1998 calendar year and data pertaining to the early part of that year become available, the degree of uncertainty faced by the forecasters diminishes.

Evidence of such convergence is readily available in panels of forecasting data. For example, Figure 1 contains the spread between eight unemployment forecasts and the mean forecast across forecast horizons for the Blue Chip data (to be described later) for 1980-1983. Forecast horizons are indexed in ascending order, so that higher index numbers are associated with shorter forecast horizons. There is considerable volatility in the data, but it is clear in all years that as the forecast horizon decreases, the distances between individual forecasts and the mean narrows.

Our objective here is to give an empirical representation to a panel of forecasters that captures the possible evolution to consensus. To this end, the reduction in the differences between competing forecasters is modelled with a finite-horizon decay process. A parametric model is used to analyse and interpret the path of forecasts towards a consensus, with a relatively small sample, in a parsimonious model. This framework also provides a direct way to test some of the restrictions of the model. One objective of the analysis is to discover whether forecast differences are predicted to decay to zero (that is, reach an endpoint) at the same point in time across a panel of forecasters. This has important implications for studies of forecaster behaviour since differences in the decay process may imply that forecasters seek to forecast different official announcements or definitions of the macroeconomic variable being forecast.

This paper is organized as follows. Section 2 develops the statistical framework, while section 3 outlines related research. Section 4 contains the estimation and results for the largest balanced panel in the Blue Chip survey of Forecasters. The analysis is extended to evaluate a larger, unbalanced panel in section 5. Section 6 then concludes.

2. Statistical Model

Recently Gregory, Smith and Yetman (2001) defined forecasters to be in consensus if their predictions are insignificantly different from the sample mean. Although many practitioners also define consensus as the sample average of forecasts at any point in time, Gregory, Smith and Yetman (2001) develop a framework to test formally (at annual frequencies) whether it is appropriate to interpret the mean as actually representing a consensus. A drawback to their approach is that each horizon is estimated and tested

separately. Overall conclusions must be drawn by connecting-the-dots across the horizons. In this study, we also treat consensus as a convergence to mean but attempt to model parametrically the time-path of this evolution by jointly using the data elements from the cross-section of forecasters, the forecast horizons, and the annual time series at each of these horizons.

In the set-up we propose, the distance between a panel of forecasts and its mean is used as an indicator of the degree of consensus among forecasters. Changes in the magnitude of the distance are used to provide insights on movements in a panel of forecasters toward, or away from, consensus along the forecast horizon line. We estimate the point at which forecasts converge, and therefore we are able to identify forecasters who are unlikely to reach consensus.

The following notation will be used. There are I forecasters, denoted by i , H forecast horizons, denoted by h , and T time periods, denoted by t . Initially we consider a complete panel where all forecasters participate for all h and t . A typical forecast is therefore $x(i, h, t)$, and the mean forecast across all forecasters excluding forecaster i for a given time period and forecast horizon is $\bar{x}(-i, h, t)$. The index h counts upwards from the first time a forecast is made, so that shorter forecast horizons are associated with larger h . The evolution of the distance between these two is modelled. The use of a mean that excludes one's own forecast reflects our interest in how each individual forecaster differs from other forecasters, rather than how he differs from the panel of forecasters as a whole. The mean is taken from the complete Blue Chip panel containing roughly 50 forecasters, so the difference in using either definition of the mean is small.

The parametric model is a finite-horizon decay process. The dynamics of the first two moments of $x(i, h, t) - \bar{x}(-i, h, t)$ are modelled as a polynomial in the forecast horizon. The model is:

$$\begin{aligned}
x(i, h, t) - \bar{x}(-i, h, t) &= \sum_{s=1}^S \mu_s(i) \text{Max}[(\delta(i) - h)^s, 0] + \varepsilon(i, h, t) \\
\text{Var}[\varepsilon(i, h, t)] &= \sum_{r=1}^R \nu_r(i) \text{Max}[(\delta(i) - h)^r, 0] \\
\text{Cov}[\varepsilon(i, h, t), \varepsilon(j, h, t)] &= \xi(i, j) \text{Max}[\text{Min}[\delta(i), \delta(j)] - h, 0]
\end{aligned} \tag{1}$$

The indexes S and R indicate the order of the polynomial in terms of $(\delta(i) - h)$ for the mean and variance respectively. In the simplest case of $S = R = 1$, $\mu(i)$ measures the rate at which the mean forecast of forecaster i decays towards the consensus, $\nu(i)$ measures the rate at which the variance of forecaster i decays towards zero, and $\xi(i, j)$ measures the rate at which the covariance between forecasters' i and j approaches zero.

It is clear that forecasters do not continue to forecast any given variable indefinitely into the future, so it makes sense to model forecast evolution with a finite-horizon decay process, rather than an exponential decay function. An end-point restriction is therefore incorporated into the model: at $\delta(i)$, which is to be estimated, forecasts by forecaster i are identically equal to the mean of all other forecasters, and the variance and covariance with each other forecaster are zero. That is, $\delta(i)$ measures the point at which forecasts by forecaster i are equal to the mean $\bar{x}(h, t)$. A 'large' $\delta(i)$ would suggest that forecaster i takes a very long time before agreeing with other forecasters. This could result, for example, if forecasters disagree on the definition of the variable being forecast or the specific 'official'

announcement of the series being forecast. As we will argue later, this is particularly relevant for GDP series that are subject to frequent revisions.

The model describes a dynamic time path for forecasters of the evolution of consensus. At $h = \delta(i)$, the variances and covariances are constrained to be zero, so that the dependent variable degenerates to zero. Note that while the forecast horizon is a discrete variable, $\delta(i)$ may take on any real value.

Model (1) imposes a certain amount of structure to the evolution of consensus, and some of these restrictions may be tested. As in Gregory, Smith and Yetman (2001) there is the assumption that the parameters are time-invariant with respect to index t . While in principle this restriction could be tested, our time dimension is much too short ($T = 18$). The model also parametrically links the time-path of forecasts across horizons using relatively few parameters. An unrestricted, though heavily parameterised, version of this model is the simple multivariate location model. For each forecaster and horizon, we would estimate a separate mean, variance, and covariance. Under some additional distributional assumptions, we test this more general model against model (1).

Another restriction we test is $\delta(i) = \delta$ for all i . This is a test that forecast differences decay to reach a consensus across our panel of forecasters at the same point in time. Note that even here, the model allows for considerable flexibility in the path to consensus for each forecaster. As will be illustrated later, one forecaster may be further from the mean than other forecasters at early horizons, but move toward the mean at a faster rate, so as to achieve consensus at the same point in time.

3. Related Research

This study is related to a number of others, each seeking to understand some aspect of the behaviour of macroeconomic forecasters. First, a number of papers show that competing professional forecasters provide a somewhat homogeneous product. For example, McNees (1979) and McNees and Ries (1983) argue that the differences between macroeconomic forecasts are small, while Batchelor (1990) demonstrates that the differences are indistinguishable from random noise. Spiro (1989) and Zarnowitz (1984) also show that forecasts are clustered very closely together. As has been pointed out by several authors, forecasting is a competitive industry with few barriers to entry; so inferior forecasters are quickly driven from business. Batchelor and Dua (1990) do find that some forecasters seek to distinguish their forecasts from each other, taking a position of optimist or pessimist within a panel of forecasters, although the magnitude of this effect is very small relative to noise present in the forecasts. Lamont (2002) finds that as forecasters become older and more established, they produce more radical forecasts that are generally less accurate. Campbell and Murphy (1996) and others also point out that forecasts are so similar that often the range of forecasts does not include the actual data when published. This is also true in the panel used here. While Rich, Raymond, and Butler (1992) find that forecast dispersion is generally correlated with uncertainty in the actual data, it is imperfectly so. The range of forecasts underestimates the degree of uncertainty facing forecasters, sometimes substantially. It follows that the degree of consensus, or agreement, found in panels of macroeconomic forecasts does not necessarily correspond to the level of information available to forecasters, or indeed the amount of uncertainty facing those forecasters.

Second, a small number of authors have sought to define what is meant by the term “consensus.” In forecasting parlance, consensus commonly referred to the mean of a panel of forecasters, as opposed to agreement among forecasters. Zarnowitz and Lambros (1987) suggest an alternative definition. Using probability forecasts, they define consensus in terms of the degree of agreement across forecasters in their point predictions, and uncertainty in terms of the diffuseness of the corresponding probability distributions. They find that the variability in point forecasts understates the degree of uncertainty as measured by the probability forecast distributions. Schnader and Stekler (1991) offered yet another definition: that consensus exists if the distribution of forecasts is symmetric, and at least as peaked as the normal distribution. Finally, Gregory, Smith and Yetman (2001) suggest a definition that can be easily tested: consensus is present when forecasters place identical weights on a common latent variable and all forecasts are insignificantly different from the sample mean.

Third, other studies have looked at the behaviour of forecasters as they update their predictions. Bachelor and Dua (1991) find that for most variables and forecasters, forecast errors diminish as the horizon shortens. Ehrbeck and Waldmann (1996) found that forecasters who make large revisions, and those who are far from the average of lagged forecasts, have large forecast errors relative to other forecasters. Davies and Lahiri (1995) provide a method for estimating forecaster behaviour that also uses all three dimensions of a data panel such as the Blue Chip. Their model is given by

$$x(i, h, t) = A(t) - \phi(i) - \lambda(h, t) + \varepsilon(i, h, t),$$

where $A(t)$ represents the realised data, $\phi(i)$ represents the bias of forecaster i , and $\lambda(h, t)$

represents the cumulative effect of shocks that are common to all forecasters. Using their framework, it is straightforward to show that our dependent variable would reduce to

$$x(i, h, t) - \bar{x}(-i, h, t) = -[\phi(i) - \bar{\phi}(-i)] - [\varepsilon(i, h, t) - \bar{\varepsilon}(-i, h, t)]. \quad (2)$$

That is, deviations from consensus are made up of a constant (the difference between the bias of forecaster i and the average bias of other forecasters), and a white-noise error term that they assume has constant variance. Therefore while Davies and Lahiri's (1995) framework incorporates convergence of forecasts towards the actual data (and so is useful for testing for forecast rationality), equation (2) excludes the possibility of convergence across forecasters towards a common forecast as the horizon shortens, as is observed in Figure 1. In contrast, equation (1) allows for variation across horizons in both the constant and the variance of the error.

The contribution of this approach is that it permits the study of the dynamic time-path to consensus, and permits a structure that differs across variables. Our empirical results suggest flexibility in the latter is essential. The structural model (1) assumes an eventual narrowing of both the distance of forecasts from the mean (a measure of consensus), and the variance around that mean (a measure of the adherence to that mean). The model postulates a degenerative distribution at some future forecast horizon so that each forecaster has identical forecasts. It seems reasonable to suppose that this might be 'close' to the actual realized value on occasion. One could generalize this model to constrain the convergence (at some distant horizon) to be 'near,' or identical to, the realized value. However, for our model, it is entirely possible for forecasters to arrive at identical forecasts that need not be equal to the actual. We think it would be worthwhile to

simultaneously explore the relationship among forecast accuracy, optimal pooling, and consensus at different forecasting horizons in a flexible dynamic specification analogous to equations (1). We leave this for future research.

Finally, McNees (1986,1992) showed that forecasts are on average closer to initial announcements than later revisions of realized data. Keane and Runkle (1990) therefore argued that forecasters seek to forecast initial announcements of realized data.

4. Balanced Blue Chip Panel

The Blue Chip survey of forecasters has been conducted each month since July 1976, originally by Robert Eggert and since 1997 by Randall Moore. Professional economic forecasters are surveyed by telephone at the start of each month. The results of the survey, including the average forecasts, are circulated around the tenth day of the month and are published in *Blue Chip Economic Indicators*.

We first estimate our model on the largest complete panel available in the Blue Chip survey. This includes the forecasts of 8 forecasters made from July of the year prior to that being forecast until May of the year being forecast, from 1979 until 1996. The variables studied are forecasts of the average annual unemployment rate, and annual gross growth rates of real GDP, nominal GDP, and the GDP deflator. The forecasters and their classifications are given in Table 1.

Model (1) is estimated by maximizing the quasi-log-likelihood function

$$L = \sum_j \sum_t \left[-\frac{1}{2} \ln |\Omega^{ht}| - \frac{1}{2} \varepsilon'_{ht} (\Omega^{ht})^{-1} \varepsilon_{ht} \right],$$

$$\varepsilon_{ht} = [\varepsilon(1, h, t), \varepsilon(2, h, t), \dots, \varepsilon(I, h, t)]',$$

$$(\Omega^{ht})_{mn} = \begin{cases} Var[\varepsilon(m, h, t)], & \text{if } m = n \\ Cov[\varepsilon(m, h, t), \varepsilon(n, h, t)], & \text{if } m \neq n. \end{cases} \quad (3).$$

The Hessian is then used to calculate standard errors in the usual way.

Notice that we can relax the restrictions in (1) for each horizon h , since the unrestricted mean in model (1) may be estimated from the individual sample mean (a simple location model), and these estimates used to estimate the variance and covariances. With these we can evaluate the unrestricted log-likelihood. For instance, for 8 forecasters the unrestricted model has 484 parameters (a symmetric (8×8) covariance matrix at each horizon, plus the mean value of $(x(i, h, t) - \bar{x}(-i, h, t))$ for each i and h). By focusing initially on a balanced panel, we can calculate this unrestricted likelihood without accounting for missing observations. Also, by withholding data at some forecast horizons, it is possible to provide a check on the model using out-of-sample prediction. In section 5, we consider an unbalanced panel to study situations in which some forecasters are not always present.

Equation (3) is interpreted as a quasi-log-likelihood, since it is very unlikely that the forecast errors are normally distributed. We also calculate likelihood ratio tests with test statistics located in the central χ^2 distribution. Again the finite sample properties of these tests may not be well approximated by this large sample test. We conducted some simple Monte Carlo experiments calibrated to the unemployment rate example (balanced panel) to check the quality of the estimates and finite sample properties of some of the tests. For the most part estimates were reasonably close, but the likelihood ratio tests conducted were biased away from the null hypothesis. Thus models containing higher order polynomials (larger S, R) were favoured by this testing procedure.

Table 2 contains a variety of tests for various specifications for unemployment rate,

real output, nominal output, and the output deflator. We begin with a detailed discussion of the unemployment rate results followed by more brief comments on the other series. Several different versions of model (1) were estimated: $S \in \{1, 2, 3\}$, and $R \in \{1, 2\}$. Including non-linear terms in the variances added little to the explanatory power of the model, so they were subsequently dropped. The quadratic and cubic representations of the parametric model appear to fit well; neither can be rejected against the unrestricted (location) model at conventional significance levels (see Table 2).

Within the class of models in (1), we test different variants of this model ($S \in \{1, 2, 3\}$). The linear model can be rejected against the quadratic, while the quadratic is only just rejected at the 5% level (see Table 2). Figure 2 contains plots of fitted values of the dependent variable for the first 11 horizons and projections for horizons beyond that for both the unrestricted and quadratic models. These plots nicely explain our findings. While there is considerable volatility in the unrestricted unemployment rate estimates, the path to convergence is concave for most forecasters so that the quadratic model provides a good fit. Exceptions to this include the Conference Board and Cahners, for whom an s-shaped curve fits substantially better than a concave function. For the panel of forecasters as a whole, a quadratic model appears adequate.

A full set of estimation results for the quadratic model is given in Table 3. Relative to the standard error, the estimates would appear to be fairly precise and illustrate the substantial variation in the forecasting behaviour of individual forecasters. For example, Chase Manhattan Bank is furthest from the mean for early horizons, but moves towards it at a faster rate than other forecasters. While the decay path decreases monotonically

for most forecasters, it increases over early horizons for Eggert Economic Enterprises. The estimated decay path for Chase Manhattan Bank is almost linear, while that for Bankers Trust is highly concave. Further, while some forecasters display systematic positive or negative forecast differences, UCLA forecasts are lower than the mean in early forecasts, but higher than the mean later on, before converging. On the other hand, Prudential Insurance forecasts display the opposite pattern. The same may be said of estimates related to the second moment. For example, the fitted variance for Cahners forecasts (not shown) are over twice as large as those for Dupont at every horizon, but decline at a faster rate.

Despite the fact that there is substantial variation in the decay process for individual forecasters, the point estimates of convergence are quite close: the estimated $\delta(i)$ varies from 11.9 months for Bankers Trust to 16.1 months for UCLA. The null hypothesis that the convergence point is common to all forecasters cannot be rejected at the 1% level (see Table 2). Further, from our Monte Carlo experiment the approximate true size for this test statistic is 0.04. At least for the unemployment rate, the evidence points to all the forecasters in our balanced panel sharing a common convergence point.

To analyse the predictive power of the model, out-of-sample forecasting is conducted for horizons 12 and 13 of years 1991 through 1996 (these additional horizons were introduced in 1991). In-sample residuals for the first 11 horizons, and out-of-sample forecast errors for horizons 12 and 13, are given in Figure 3. It is clear that for most years and forecasters, the magnitude and direction of the forecast errors is in line with residuals from the earlier horizons and again suggests the model does a reasonable job out-of-sample too.

Turning to the three output series, for real and nominal output growth, the restricted model appears to fit the data well, and cannot be rejected at conventional levels against the unrestricted model (see Table 2). For the output deflator, the fit is substantially worse, although both the quadratic and cubic restricted models cannot be rejected against the unrestricted model. A cubic term in $(\delta(i) - h)$ adds little to the explanatory power over the quadratic model. The point of decay varies across forecasters and series, from 12 to 22 months, with the exception of the nominal and real output forecasts of the UCLA, for whom a very large and imprecise estimate is obtained. In contrast to the unemployment forecasts, the point of decay is significantly different across forecasters, even excluding UCLA (an outlier for these variables) from the test. For these series, forecasters do not all converge to a consensus at the same point in time.

There are several possible explanations for differences in the end point to the decay process across forecasts and the lack of an end point for UCLA forecasts. As we have suggested, strategic interactions between forecasters may influence the timing and evolution of a consensus. Perhaps, more importantly with respect to the output measures, forecasters disagree on the definition of the series being forecast, for example by seeking to forecast different announcements of the ‘official’ data. The official statistics that are the companion to our output forecasts are revised up to 7 times and published by the United States Department of Commerce in the *Survey of Current Business*. The realized data first becomes available in the middle of December of the year being forecast, and incorporates projections of the level of economic activity for most of the final quarter. It is then revised in the middle of January, February, and March, and then several additional times after

that. If all the forecasters are seeking to forecast the initial publication of each of these series, then one would expect to obtain estimates of $\delta(i)$ no greater than 18 (the value of h when the GDP figures are first published). If all forecasters are seeking to forecast the series after it is no longer subject to revision, estimates of $\delta(i)$ may indeed be much larger than this.

This raises an interesting possibility. If forecasters seek to forecast different announcements of the official statistics, one would expect to find wide dispersion in the estimated $\delta(i)$'s with some forecasters never converging to the consensus. This would be possible as long as there is a forecastable component in the revisions. Since initial announcements are little more than forecasts produced by the official statistical agency that incorporate predictions of economic activity in the final months of the period being measured, knowledge of the process by which official statistics are produced, in particular the information that is incorporated in early announcements versus that which is included in later measures, would need to be incorporated into the model. However, such an endeavour is beyond the scope of this paper.

In contrast to the output series, there are fewer complications for the unemployment data used in this study. There is typically a lag of less than one month between the end of the calendar year and the publication of the annual unemployment rate, and after that the rate is subject to few, if any, revisions. This may explain our finding of a common end point across unemployment forecasts.

5. Unbalanced Blue Chip Panel

Notice in model (1) that there is no special need to have a complete panel since we can

estimate the shape of the decay function with missing data. In fact, we could have entire horizons missing and simply use the polynomials to interpolate over the missing horizons. However, we estimate model (1) on an unbalanced panel incorporating the most frequent forecasters over the history of the Blue Chip survey for the 11 horizons considered above. We limit our analysis to an ‘almost’ complete panel for two reasons. First a forecaster’s decision to participate or not may bias our results. For example, if a forecaster chooses to release to survey collectors only those forecasts that are close to the forecasts of others, she may appear to mimic the consensus of others when in reality her forecasts do not. This applies especially to the earliest and latest forecast horizons for which the Blue Chip panel is collected, as some forecasters simply do not provide forecasts for specific horizons at either end. Second, the degrees of freedom decline sharply as the number of forecasters in the panel increases and estimation/convergence problems abound when we include forecasters with sparse data. As a result we limit ourselves to a panel of 14 forecasters who are each present over 90 percent of the time; the next most frequent forecaster is present only 79 percent of the time. This expanded panel commences in February 1977 and ends in May 1997.

In practical terms, estimating our model with an unbalanced panel requires that the error corresponding with each missing observation be censored so that it does not contribute to the likelihood (the observation is replaced by the conditional expectation for the mean, variance and covariance). Estimating the quadratic version of model (1) involves 144 parameters. Not surprisingly, as we add additional forecasters to the panel, a common decay point for all forecasters is rejected, with at least one forecaster for each variable fail-

ing to converge to the consensus. Point estimates of $\delta(i)$ are in some cases in excess of 140 months. The non-converging forecasters are Bank of America (Unemployment), Dupont and UCLA (real GDP), UCLA (nominal GDP), and National City Bank of Cleveland (GDP deflator).

Our framework can also be used to test hypotheses on subgroups of forecasters to see whether for some subsets their behaviour is different. For example, does convergence in one variable imply convergence in others? Four of the forecasters seem largely in agreement (Canhars Economics, the Conference Board, Eggert Economic Enterprises and Prudential Insurance) and were selected for separate study. For these forecasters, we cannot reject a common decay point for unemployment, real output, or nominal output on the unbalanced panel, while we can reject a common decay point for the output deflator (significance levels of 0.26, 0.08, 0.23, and 0.001 respectively). Of course, selecting the forecasters on the basis of observed characteristics invalidates the testing procedure. However, our goal here is to show how the procedure can be applied in other situations where the investigator may wish to compare forecasters using criteria not related to the data. For instance, one could test whether bank forecasters are different from other private sector forecasts.

6. Conclusions

When professional forecasters repeatedly forecast macroeconomic variables, their forecasts converge over time perhaps towards a consensus. We have analysed panels of Blue Chip data using a simple polynomial decay function that incorporates this feature, yet permits flexibility in the decay path. We have shown that this empirical representation of the data cannot be rejected against a more general location model.

Using the largest complete panel of Blue Chip forecasts, unemployment forecasts share a common end point to the decay process, while the end point for GDP forecasts differs significantly across forecasters. We argued that one plausible explanation for this is that while there is agreement among forecasters about the definition of unemployment being forecast, frequent revisions in output series may result in forecasters seeking to forecast different announcements of the official data. If this explanation is correct, care must be taken when using a panel of forecasts. An implicit assumption in combining forecasts is that forecast differences reflect either differences in the information sets of forecasters, or different methods employed by forecasters. The alternative possibility suggested here is that forecasters differ in what they are trying to forecast (in terms of the announcement date), even if they define the variable in an identical fashion.

When we estimate our model using a larger, unbalanced panel, a common convergence point can be rejected for all variables considered. The framework is also useful in testing for the emergence of consensus on a subset of forecasters.

Table 1. Blue Chip Forecasters.

Forecaster	Title	Classification
Balanced Panel		
1	Bankers Trust Co.	Bank
2	Chase Manhattan Bank	Bank
3	Conference Board	Industry Association
4	Cahners Economics	Financial Publication
5	Eggert Economic Enterprises, Inc.	Financial Publication
6	UCLA Business Forecasting	Econometric Modellers
7	Prudential Insurance Co.	Insurance Company
8	Dupont	Industrial Corporation
Unbalanced Panel		
9	Bank of America	Bank
10	Brown Brothers Harriman	Bank
11	National City Bank of Cleveland	Bank
12	U.S. Chamber of Commerce	Industry Association
13	JP Morgan	Bank
14	U.S. Trust Co.	Bank

Note: The classifications are taken from Laster *et al* (1997).

Table 2. Test Statistics.
Balanced Panel.

H_O	H_A	Statistic	d.f.	χ^2 location
Unemployment				
Linear	Unrestricted	424.1	432	0.5983
Quadratic	Unrestricted	402.5	424	0.7663
Cubic	Unrestricted	386.8	416	0.8445
Linear	Quadratic	21.5	8	0.0059
Linear	Cubic	37.2	16	0.0019
Quadratic	Cubic	15.7	8	0.0464
$\gamma(i) = \gamma \quad \forall \quad i$	$\gamma(i) \neq \gamma$	16.5	7	0.0208
Real Output				
Linear	Unrestricted	391.0	432	0.9221
Quadratic	Unrestricted	370.6	424	0.9710
Cubic	Unrestricted	362.0	416	0.9736
Linear	Quadratic	20.4	8	0.0089
Linear	Cubic	29.0	16	0.0240
Quadratic	Cubic	8.6	8	0.3782
$\gamma(i) = \gamma \quad \forall \quad i$	$\gamma(i) \neq \gamma$	41.5	7	0.0000
Nominal Output				
Linear	Unrestricted	400.3	432	0.8606
Quadratic	Unrestricted	383.6	424	0.9208
Cubic	Unrestricted	378.2	416	0.9079
Linear	Quadratic	16.7	8	0.0334
Linear	Cubic	22.1	16	0.1406
Quadratic	Cubic	5.4	8	0.7160
$\gamma(i) = \gamma \quad \forall \quad i$	$\gamma(i) \neq \gamma$	32.9	7	0.0000
Output Deflator				
Linear	Unrestricted	492.9	432	0.0226
Quadratic	Unrestricted	457.6	424	0.1257
Cubic	Unrestricted	446.2	416	0.1480
Linear	Quadratic	35.3	8	0.0000
Linear	Cubic	46.7	16	0.0001
Quadratic	Cubic	11.4	8	0.1800
$\gamma(i) = \gamma \quad \forall \quad i$	$\gamma(i) \neq \gamma$	21.7	7	0.0029

Note: Tests of the form $H_O : \gamma(i) = \gamma \quad \forall \quad i$ are for the quadratic model.

Table 3. Estimation Results.Unemployment, Quadratic model.

$$x(i, h, t) - \bar{x}(-i, h, t) = \mu_1(i)(\delta(i) - h) + \mu_2(i)(\delta(i) - h)^2 + \varepsilon(i, h, t)$$

$$Var[\varepsilon(i, h, t)] = \nu(i)(\delta(i) - h)$$

$$Cov[\varepsilon(i, h, t)\varepsilon(j, h, t)] = \xi(i, j)(\min(\delta(i), \delta(j)) - h)$$

Forecaster	Variable	Coefficient	Standard Error	Significance
<hr/>				
Bankers Trust	$\mu_1(1)$	-0.0351	0.0062	**
	$\mu_2(1)$	0.0019	0.0007	*
	$\nu(1)$	0.0073	0.0007	**
	$\delta(1)$	11.923	0.2544	**
Chase	$\mu_1(2)$	0.0123	0.0070	
	$\mu_2(2)$	0.0014	1.5684	
	$\nu(2)$	0.0161	0.0019	**
	$\delta(2)$	12.070	0.3400	**
Conference Board	$\mu_1(3)$	0.0096	0.0063	
	$\mu_2(3)$	-0.0001	0.0007	
	$\nu(3)$	0.0135	0.0014	**
	$\delta(3)$	12.929	0.4780	**
Cahners	$\mu_1(4)$	0.0089	0.0096	
	$\mu_2(4)$	-0.0009	0.0011	
	$\nu(4)$	0.0180	0.0015	**
	$\delta(4)$	12.864	0.4997	**
Dupont	$\mu_1(5)$	-0.0101	0.0050	*
	$\mu_2(5)$	0.0003	0.0005	
	$\nu(5)$	0.0073	0.0007	**
	$\delta(5)$	14.431	0.7199	**
Prudential	$\mu_1(6)$	-0.0168	0.0051	**
	$\mu_2(6)$	0.0015	0.0005	**
	$\nu(6)$	0.0128	0.0017	**
	$\delta(6)$	13.900	0.7630	**
Eggert	$\mu_1(7)$	-0.0268	0.0067	**
	$\mu_2(7)$	0.0014	0.0007	*
	$\nu(7)$	0.0110	0.0011	**
	$\delta(7)$	13.868	0.6163	**
UCLA	$\mu_1(8)$	0.0281	0.0083	**
	$\mu_2(8)$	-0.0022	0.0009	*
	$\nu(8)$	0.0139	0.0029	**
	$\delta(8)$	16.128	1.9349	**

Note: * and ** indicate significance at the 5% and 1% level respectively.

References

- Batchelor, Roy A. (1990) All Forecasters Are Equal. *Journal of Business and Economic Statistics* 8, 143-144.
- Batchelor, Roy A. and Pami Dua (1990) Product Differentiation in the Economic Forecasting Industry. *International Journal of Forecasting* 6, 311-316.
- Batchelor, Roy A. and Pami Dua (1991) Blue Chip Rationality Forecasts. *Journal of Money, Credit, and Banking* 23(4), 692-705.
- Campbell, Bryan and Steve Murphy (1996) Who Forecasts Best? An Empirical Analysis of Canadian Private-Sector Forecasters. *Canadian Business Economics*, 4(3), 3-17.
- Davies, Anthony and Kajal Lahiri (1995) A New Framework for Analyzing Survey Forecasts Using Three-Dimensional Panel Data. *Journal of Econometrics* 68, 205-227.
- Ehrbeck, Tilman and Robert Waldmann (1996) Why are Professional Forecasters Biased? Agency Versus Behavioral Explanations. *Quarterly Journal of Economics* 111(1), 21-40.
- Gregory, Allan W., Gregor W. Smith, and James Yetman (2001) Testing for Forecast Consensus. *Journal of Business and Economic Statistics* 19(1), 34-43.
- Keane, Michael P., and Runkle, David E. (1990) Testing the Rationality of Price Forecasts: New Evidence From Panel Data. *American Economic Review* 80, 714-735.
- Lamont (2002) Macroeconomic Forecasts and Microeconomic Forecasters. *Journal of Economic Behavior and Organization* 48, 265-280.
- Laster, David, Bennett, Paul, and Geoum, In Sun (1999) Rational Bias in Macroeconomic Forecasts. *Quarterly Journal of Economics* 114(1), 293-318.
- McNees, Stephen K. (1979) The Forecasting Record of the 1970s. *New England Economic Review* Sep/Oct, 33-53.
- McNees, Stephen K. and John Ries (1983) The Track Record of Macroeconomic Forecasts. *New England Economic Review* Nov/Dec, 5-18.
- McNees, Stephen K. (1986) Estimating GNP: the trade off between Timeliness and Accuracy. *New England Economic Review* Jan/Feb, 3-10.
- McNees, Stephen K. (1992) How Large are Economic Forecast Errors? *New England Economic Review* Jul/Aug, 25-33.
- Rich, R.W., J.E. Raymond, and J.S. Butler (1992) The Relationship between Forecast Dispersion and Forecast Uncertainty: Evidence from a Survey Data-Arch Model. *Journal of Applied Econometrics* 7, 131-148.
- Schnader, M.H. and Stekler, H.O. (1991) Do Consensus Forecasts Exist? *International Journal of Forecasting* 7, 165-170.

- Spiro, Peter S. (1989) Improving a Group Forecast by Removing the Conservative Bias in its Components. *International Journal of Forecasting* 5(1), 127-131.
- Zarnowitz, Victor (1984) The Accuracy of Individual and Group Forecasts from Business Outlook Surveys. *Journal of Forecasting* 3, 11-23.
- Zarnowitz, Victor and L.A. Lambros (1987) Consensus and Uncertainty in Economic Prediction. *Journal of Political Economy* 95, 591-621.

Figure 1. Unemployment Forecasts
Distance from Mean

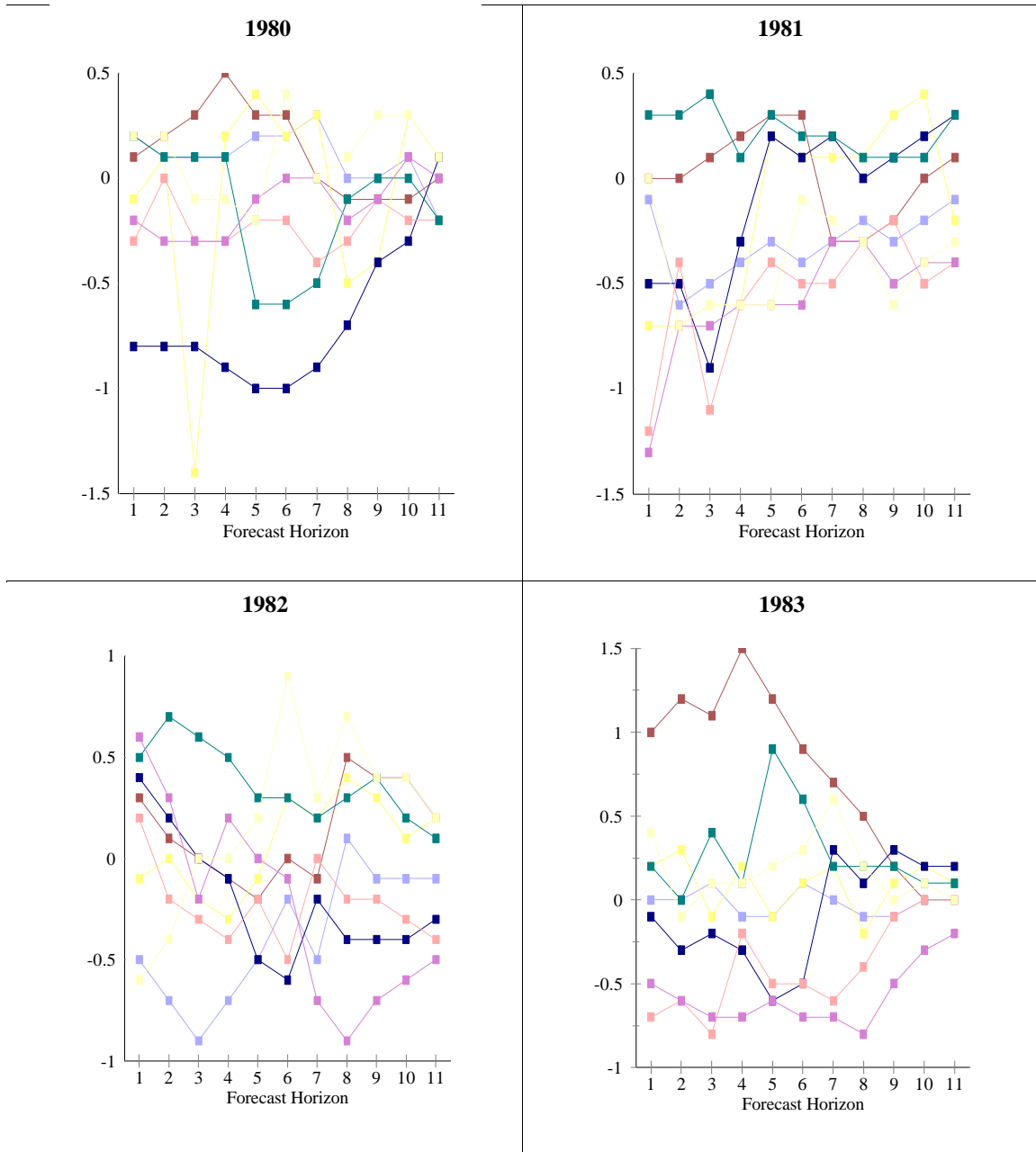


Figure 2. Unemployment Estimates
Quadratic and Unrestricted Models

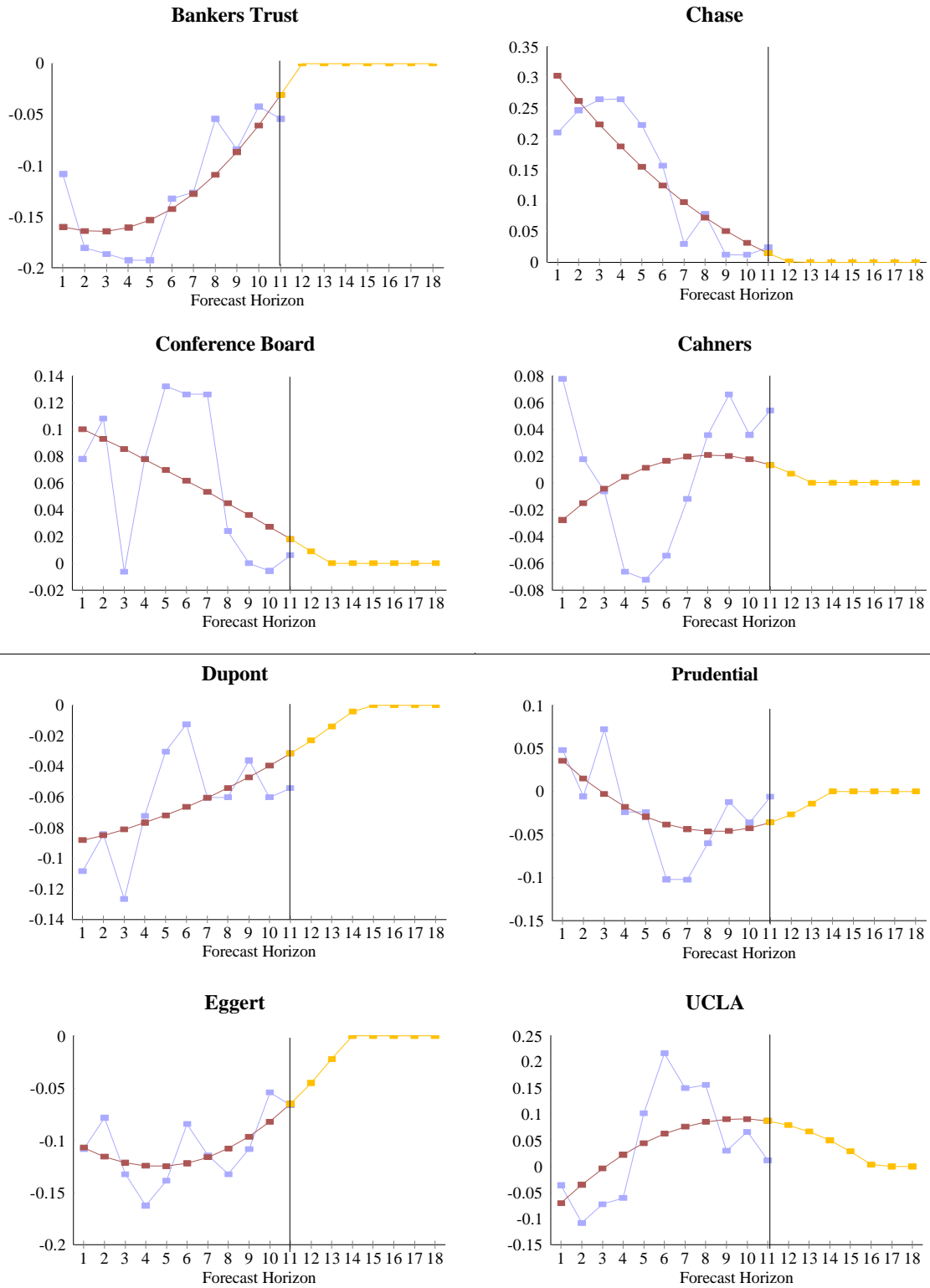


Figure 3. Unemployment Residuals and Prediction Errors

